DOCUMENT RESUME

ED 262 386 CS 008 176

AUTHOR Spiegel, Glenn; Campbell, John J.
TITLE Measuring Readability with a Computer: What We Can Learn.
PUB DATE May 85
NOTE 18p.; Paper presented at the Meeting of the UCLA Conference on "Computers and Writing: New Directions in Teaching and Research" (Los Angeles, CA, May 4-5, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Assisted Testing; *Computer Software; Higher Education; Measurement Techniques; *Readability; *Readability Formulas; *Reader Response; *Reading Comprehension; Test Validity; *Word Processing; Writing (Composition)
IDENTIFIERS *Audience Awareness

ABSTRACT

The Flesch readability index yields meaningful information about the responses of readers to texts. Because the formula is so simple, a group of English teachers wrote a program in BASIC that would count some obvious surface features of a text and calculate Flesch scores. Among the programing problems encountered were counting words (taking into account numbers, acronyms, and abbreviations) and counting the number of syllables in a word (English has no regular rules for reliably dividing words into syllables). Using only this readability program will not ensure improved comprehension because the act of reading and comprehending involves so many interrelated factors. However, the program can be used to test the readability of government manuals, orders, instructions, and so on. It can also be used to alert writers to revise their prose according to the reading level of the intended audience and to serve as a taking-off point for a college classroom discussion of audience. An advantage of the readability program is that students can run it themselves. For those who use word processors, a readability program will probably end up as one more utility program--like a word counter or spelling checker--that quickly provides potentially useful information about one's prose. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John J. Campbell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED262386

Measuring Readability with a Computer

What We Can Learn

by

Glenn Spiegel, US General Accounting Office

John J. Campbell, Howard University

What can computers tell us about writing that is meaningful in human terms, that goes beyond merely counting and tabulating? Most of us looking at computer applications to writing seem to be up against a corollary of Murphy's law that states, "The more meaningful the question we ask, the less likely we are to get an answer from the computer."

Readability measurement seems, at first glance to be an area where something a computer does well--counting--can yield meaningful information about human activities--the response of readers to texts. And so we set out to do something we thought was simple--write a program in BASIC that would run on our microcomputers, count some obvious surface features of a text, do a little arithmetic, and calculate a commonly used index of readability, the Flesch index. This index, we felt, could tell a writer useful things about how easily his or her prose (or someone else's) will be understood. Readability scores, after all, are usefully applied to a range of things from children's books to procedures for nuclear reactor safety.

CS008 176

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Of course there was no escaping Murphy's Law. Even "simple" surface features of a text make severe demands on a computer's capabilities, and a readability score turns out not to be quite as meaningful as one would hope.

Measuring Readability

Readability is measured by reassuringly concrete methods. Readers are given a text to read and are then tested on how much they understand. Two popular testing methods are simply asking questions about the passages and the "cloze" procedure, in which people are given a text with words deleted and are asked to guess what the deleted words are.[1] Both these methods test the actual response of readers to actual texts.

Readability formulas are developed by finding numerical measures of surface features that correlate well with the scores obtained from actual readers. In a pioneering work, Gray and Leary[2] rewrote standard texts in a variety of ways to test for the effects of 44 different style variables--such as sentence length, whether verbs are active or passive, and number of pronouns--on readability. (They recognized that content, format, and organization also affected readability, but felt that only style could be quantified and tested objectively.) They found that 20 of

---

[1]J.R. Bormuth, "Readability: A New Approach," Reading Research Quarterly, vol. 1 (1966), pp. 79-132.

[2]W.S. Gray and B.E. Leary, What Makes a Book Readable? (Chicago, Univ. of Chicago Press, 1935).

their variables were significant, but that five accounted for most of the variance. Increasing the following made prose harder to read:

- number of hard words used
- number of different words used
- average sentence length
- number of prepositional phrases

Increasing the following made prose easier to read:

- number of personal pronouns

The first four of these are rough measures of diction and syntactic complexity, while the fifth seems to relate to reader interest.

Rudolf Flesch, another pioneer, developed a simpler formula that takes into account vocabulary and syntax and that correlates well with tests on readers.[2] Flesch's formula calls for counting the words and syllables in a sample of text. (He cautions that this does not mean that word and sentence length are the only determinants of readability, only that they can serve as indicators.)

The formula gives a numerical score, RE (for Reading Ease), determined as follows:

RE = 206.835 - .846*(syllables/100 wds.) - 1.015*(av. wds./sent.)

(Note the six-digit precision of the first coefficient; we'll return to this later.)

---

[2]W.S. Gray and B.E. Leary, What Makes a Book Readable? (Chicago, Univ. of Chicago Press, 1935).

Flesch suggests interpreting the results by the following table:

0-30 Graduate or specialized knowledge required.
30-50 College level
50-60 High School level
60-70 Eighth-grade level
70-80 Seventh-grade level
80-100 Sixth-grade or lower level

In the second column, the information that really counts, we now have one-digit precision, and even that is a little questionable. We can put the number 8 on a certain grade level, but that really gives us only a probability that any given student will be able to read the material. How many eighth graders read at exactly the eighth-grade level?

Another popular formula of this type was developed by Robert Gunning, who used word and sentence length in a slightly different way to produce what he calls the "fog index."[4] It is slightly less reliable than the Flesch score but has enjoyed considerable success among those who enjoy flogging bureaucratic prose but don't necessarily know anything about writing. (A danger we will return to later is the danger of putting numbers into the hands of people who don't understand the concepts behind them.)

Whatever the accuracy, though, a Flesch score tells us experimentally validated truth about something of genuine interest--how hard a text is to understand. Furthermore, the

---

[4]Gunning, Robert, The Techniques of Clear Writing (New York: McGraw-Hill, 1952).

formula is so simple that even an English teacher can write a program to calculate these values. Well, maybe.

Writing the Program: Ambiguity and Compromise

The first programming problems came up in what should be the simplest task--counting the words. What does one do about numbers, acronyms, and abbreviations? Should an acronym be counted as a word for each letter? The first compromise was to count any number or acronym as one word, even though we know they impair readability.

The next problems came from the ambiguity of written English surface structures. Sentences would not seem to be hard to count; after all they are well marked by initial capitalization and terminal punctuation. But a computer needs to have every detail in place, and, after all, a period does not always mark a sentence end, and neither does a question mark. (It may be part of an embedded quotation.) A period followed by two spaces does always mark a sentence end (at least in accurately typed text) but, if it ends a paragraph, it will be followed, not by two blank spaces, but by a carriage return, as would an abbreviation that happens to come at the end of a line. These problems can be solved (except maybe for sentences with embedded quotations) but some inaccuracy creeps in. The whole problem is much less trying though, if one keeps in mind that the answer will have only one significant digit.

The worst problem is with syllables. English has no regular rules that will reliably divide words into syllables. Typesetters use complex computer algorithms coupled with tables of exceptions to

determine word breaks,[5] and these have to be checked by human editors. On a mainframe, one could, of course, include in one's program a table of all the words the program might encounter with the number of syllables each has, but that's out of the question for a micro, and setting up the table was too boring even to contemplate.

Of course people have trouble with syllables, too. Phoneticians don't agree on what defines a syllable, nor do speakers of different dialects of English agree on how to pronounce words. Does "idea" have two syllables or three? It depends on who you ask. At an early stage, we considered getting graduate assistants to do some syllable counts for us to validate our program, but we discovered that they weren't very good at recognizing and counting syllables either. One of us found that, despite his Ph.D. in English, he was unable to get reproducible results counting syllables in a 200-word sample passage. Fortunately, we were saved again by the one-digit precision of the Flesch score. If we count letters and divide by 3.0, we get a value that will rarely vary more than about ten percent from the counted number of syllables for the type of prose (government reports) we are mainly working with.

So, finally, having been rescued from the computer's need for exact specification by the imprecision of human life, we finally had a program that calculated Flesch scores. We were now ready to

---

[5]See, for example Donald Knuth, Tex and Metafont:New Directions in Computer Typesetting (New York: 1979).

explore just what our results could be used for.

What Do Readability Measures Tell Us About Actual Reading?

As we have seen, readability formulas take as their indicator variables two factors from the many involved in the complex act of reading to comprehend ideas. Several other factors may have more influence on comprehension than the style characteristics used to arrive at a readability score.

The act of reading and comprehending ideas from the printed page requires the interaction of the reader and the text. Characteristics of both reader and text influence comprehension. Reader characteristics include purpose, motivation, interest, knowledge of content, experience with the type of document, and ability to use an appropriate reading strategy. Text factors involve prose and non-prose influencers. Prose factors include vocabulary and syntax (the two factors Flesch's formula addresses) as well as concept load, concept density, and the organization of the ideas in the text. Non-prose factors include the reader's environment and the use of illustrations.

A reader's purpose, motivation, and interest in comprehending the ideas in the document influence greatly the amount of concentration devoted to comprehension. Consider the casual reading of a novel for pleasure. The reader is not under any external pressure to prove that comprehension has taken place, thus, he/she has the flexibility of choosing what to remember. Now suppose that the same novel is <u>assigned</u> to two college students--a business

major, and an English major. Both would be under pressure to prove comprehension. Perhaps the business major would view the reading and comprehension in the larger context of all the courses required for graduation. On the other hand, the English major would have a higher degree of purpose, motivation, and interest in attaining high comprehension. Both, of course, would wonder if it was going to be on the final, but the English major would be more motivated to get a high grade in an English course and would be likely to try to remember more of the text. Finally, consider what would happen to the students' motivation and retention if the final exam questions on the book were handed out when it was assigned. Research indicates that providing readers with questions before and during the reading act (giving the reader specific purposes and motivation) improves comprehension.[6] According to one study,[7] a reader's knowledge of the content being read is the most important variable influencing comprehension. Reading about a familiar subject involves fitting the ideas into a well-defined memory structure. The more information a reader has in his/her cognitive structure the more active the comprehension process. Familiar ideas are

---

[6]See T. Andre, "Does Answering Higher-level Questions While Reading Facilitate Productive Learning? Review of Educational Research, 49 (1983), pp. 280-318, and E. Rothkopf and B. Bloom "Effects of Interpersonal Interaction on the Instructional Value of Adjunct Questions in Learning from Written Material," Journal of Educational Psychology, vol. 61 (1970), pp. 417-422.

[7]D. Ausubel, The Psychology of Meaningful Verbal Learning (New York: Grune and Stratton, 1963)

reinforced, new ideas are placed and connected to an existing knowledge structure. However, the reader who has little or no knowledge base must begin to develop the cognitive structure while reading. In many cases, the reader is unaware of which ideas are most important and which are the supporting ideas. The reader who knows little about the subject is likely to focus on the word level of comprehension, rather than the ideas, in order to decode new terminology.

Experience with the format of the material being read also affects comprehension. A college student may have considerable experience in reading and comprehending a textbook format, for example. This experience allows the student to predict the location of important ideas in the text by using previous knowledge about how a typical textbook is organized. But the student might lack experience with another type of document--for example, work-related procedures or tax instructions--and would have to work to discover the organization of ideas before comprehending the structure of the ideas.

Finally, the reader needs to use appropriate reading strategies to comprehend ideas. Goodman[8] found that readers who scored low on a comprehension test were not aware of the specific strategies used to comprehend written text. They reported a focus on decoding the

---

[8]K. Goodman, "A Linguistic Study of Cues and Miscues in Reading," Elementary English, Oct., 1965, pp. 639-643.

individual words in the text, not in searching for the logic and structure of the ideas. Readers who scored high on the comprehension test reported using a variety of reading strategies to locate and process the ideas in the text. For example, they used skimming and scanning techniques to preview the material, they turned headings into questions in order to read for a specific purpose, and they reread to gain a clearer understanding and link ideas.

The above discussion has focused on the variables associated with the audience for a document. No readability index will tell us anything about them, nor will any rewriting of the text do anything about them. We must still consider relevant prose and non-prose features of the text, however, and we can revise these to improve comprehension.

Semantic features of a text also influence how well a reader comprehends. In particular, the complexity and density of ideas and how these ideas are organized influence comprehension. For example, Kintsch[9] and Kintsch and Van Dijk[10] have found that the density or number of relationships among ideas in a passage significantly affects readers' ability to recall information. They tested readers on passages that differed in the number, complexity, and

---

[9]W. Kintsch, The Representation of Meaning in Memory (New York: Wiley, 1974).

[10]W. Kintsch and T.A. Van Dijk, "Toward a Model of Text Comprehension and Production," Psychological Review, 85 (1968), pp. 363-394.

organization of ideas included but that had similar readability scores. Readers remembered more when the number of ideas was reduced and they were presented in a hierarchical order. Frase[11] among others, has reported better comprehension results when subjects were given passages arranged in a logical order compared to passages where the same sentences were randomly ordered.

Segmenting prose text into different content groups also improves comprehension scores. For example, Frase and Schwartz[12] took a standard paragraph text format and segmented the sentence components by different forms of indentation. The segmented text resulted in 14 to 78 percent faster responses to questions about the text. Thus, it appears that both segmentation and indentation influence comprehension.

Non-prose factors also seem to influence comprehension. Research in human factors indicates that environmental factors such as heat and light, affect a reader's comprehension. Moreover, specific features in a document--e.g. size, legibility of print, color--influence not only what a reader interacts with but also how the interaction occurs and what is gained from it.[13]

---

[11]L. Frase, "Influence of Sentence Order and Amount of Higher-Level Text Processing Upon Reproductive and Productive Memory," American Educational Research Journal, vol. 5 (1976), pp. 307-319.

[12]L. Frase and B. Schwartz, "Typographical Clues that Facilitate Comprehension," Journal of Educational Psychology, vol. 71, (1979) pp. 197-206.

[13]M.M. MacDonald-Ross, Language in Texts: The Design of Curricular Materials, in in L.S. Shulman, ed., Review of Research in Education, vol. 6 (Peacock Publishers, Inc., 1978)

The use of graphic devices has been reported to aid comprehension. MacDonald-Ross reports on several studies indicating that graphic devices such as tables, graphs, charts, illustrations, color, margin widths, and highlighting techniques (headings, subheadings, italics, underlining, etc.) have improved comprehension.[14]

Thus, the reading/comprehension act involves many interrelated factors. These factors, related to the reader and the text, can have a significant affect on the comprehension of the ideas represented by the prose. Merely applying a readability formula, which looks at only part of the language factor, will not ensure improved comprehension.

Where are readability scores used?

Children's books and school texts would seem to be the ideal field for readability measurements. For one thing, the difference among different reading levels at different ages is more universal and developmental than the differences in vocabulary and skills among adults, which may be determined by employment, reading, and so on. We can count on more homogeneity among children at a given stage of development than we can among adults. Indeed, most children's texts are tested for readability before publication. (Testing may however be done using methods other than Flesch scores,

---

[14]M.M. MacDonald-Ross, "Graphics in Text," in L.S. Shulman, ed., Review of Research in Education, vol. 5 (Peacock Publishers, Inc., 1978)

such as the Dale-Chall score, which checks to see how many words in the text come from beyond a standard vocabulary thought to be known by most fourth graders.[15]

At the other end of the spectrum, the Department of Defense uses readability testing extensively in preparing manuals, orders, instructions, and so on. The average recruit, who may have to maintain a state-of-the-art piece of electronic gear, reads at the fourth-grade level. The military takes readability measurement so seriously that the Navy, for example, maintains a list of words they expect their ratings to know and continually rechecks the words on the list through a testing program. The Naval Training Center, Orlando, FL, has developed a computer program that checks text for words not on this list and suggests synonyms from the list.[16]

One of the authors of this article helped prepare a report for the U.S. General Accounting Office on the effectiveness of automobile recall letters. The report used readability measurements of actual recall letters to show that most of the people they were addressed to would have serious difficulty in comprehending the message, since the letters were written at the graduate level or above and the average reading level of the American public is the

---

[15]E. Dale and J. Chall, "A Formula for Predicting Readability," Educational Research Bulletin, Jan. 21 and Feb. 17, 1948, pp. 11-20 and 37-54.

[16]J.P. Kincaid, J.A. Aagard, and J.W. O'Hare, Development and Test of a Computer Readability Editing System (CRES), (Orlando, FL:Training Analysis and Evaluation Group, 1980), Technical Report No. 83.

eighth grade. We validated this conclusion by informal testing of the actual automobile return rate for the original letters (17%) and for a rewritten and reformatted version (84%).

Caveats

The General Accounting Office issues 500 to 700 lengthy reports a year and is concerned about the accessibility of its findings. It would seem to be an ideal place to introduce the use of readability formulas such as the Flesch score. We have been reluctant to do so, however, for reasons already alluded to.

Many GAO staff members are accountants and social scientists by training; few have professional expertise in writing, but most feel safe dealing with numbers. For this very reason, we don't encourage them to use readability formulas. We feel this might lead to writing by numbers. We especially worry that unskilled writers might assume the formula is a rewriting rule. (Although many authorities have warned against using readability formulas as a guide for rewriting material,[17] the formulas are being used this way.)

Anyone can write short words and sentences. But the result is not necessarily good. Short sentences make prose choppy. Short words may cause clumsy explanations of things that have good long names. This paragraph is an example.

---

[17]See, for example, G.R. Klare, Readability Standards for Army-wide Publications (Fort Harrison, Ind.: U.S. Army Administrative Center, 1979).

The point we are making is that readability formulas may not even be useful in situations that seem most promising, because they don't say enough about prose that is meaningful. They are too simple and too liable to misinterpretation. No number can tell us much about prose by itself; it needs an expert to explain and apply it. It will not give an unskillful writer more expertise, it can only help a skillful writer to diagnose potential problems and fine-tune his prose.

How Do We Use Readability Formulas?

One of us regularly runs our readability program on all his prose; he can call it up from inside his word processing software. Generally, he gets Flesch scores of around college level. If this is the case, he does nothing special; he is writing for people with college degrees. For one project, however, which called for writing instructions for secretaries on typing a new document format, he found that a Flesch score of college level triggered a rewrite. He spent most of his effort in reformatting the document, however, not in changing words and sentences. He changed the segmentation and indentation, paying some attention as well to vocabulary, but little to sentence length. Merely shortening the sentences in a list of instructions is unlikely to help the secretary at the word processor who is still looking for the instructions.

How Can Readability Measurement Be Useful in the College Classroom?

Certainly one wouldn't want to tell college students, any more than accountants, to make their already underdeveloped sentences any

shorter. In fact, we think the most valuable use of readability as a concept in the college classroom is as a taking-off point for a discussion of audience, a group that is all too sadly missing from most college writing projects. A college student can safely assume that his professor can and will get through the prose of his paper somehow, no matter how bad it is. After all, the professor is a skilled reader, and it is his or her job to read student papers and even write comments on them to prove they have been understood. Writing a paper for someone with less skill and less knowledge (Dare we say less interest?) might prove a very useful challenge, if prefaced by a discussion of what really influences readers' comprehension. The assignment itself might usefully call for the student to write instructions, perhaps for a word processor, and the instructions should actually be tested on other students.

An advantage of a readability program is that students can run it themselves. The novelty of a new toy, coupled with the rapidity of the feedback may encourage students to look carefully at their writing (but beware of writing by the numbers). This would be most appropriate to upper-level students who have become sufficiently expert in some subject area to clog their prose with jargon.

A readability score may also be helpful in that it cannot be dismissed by students as a teacher's personal preference. After all, no matter what grade last semester's teacher gave him or her, a student can hardly argue with a number generated by simple counting and multiplying.

For those students (and faculty) who use word processors, a readability program will probably end up as one more utility program, like a word counter or spelling checker, that provides, quickly and easily, some potentially useful information about one's prose. Ultimately, it is not going to be a breakthrough into an area where computers can tell us something profound and meaningful about writing. What it may do is prompt us to apply our own knowledge.[18]

---

[18]We will be happy to send you a listing of our program (written in BASIC) or to transmit it to you electronically. Contact Glenn Spiegel, Writing Resources Branch, US GAO, 441 G St. NW, Room 4528, Wash. DC, 20548.